



Learning from evaluation at the Millennium Challenge Corporation

Jennifer Sturdy, Sixto Aquino & Jack Molyneaux

To cite this article: Jennifer Sturdy, Sixto Aquino & Jack Molyneaux (2014) Learning from evaluation at the Millennium Challenge Corporation, Journal of Development Effectiveness, 6:4, 436-450, DOI: [10.1080/19439342.2014.975424](https://doi.org/10.1080/19439342.2014.975424)

To link to this article: <http://dx.doi.org/10.1080/19439342.2014.975424>



© 2014 The Author(s). Published by Taylor & Francis.



Published online: 12 Nov 2014.



Submit your article to this journal [↗](#)



Article views: 1719



View related articles [↗](#)



View Crossmark data [↗](#)

Learning from evaluation at the Millennium Challenge Corporation

Jennifer Sturdy*, Sixto Aquino and Jack Molyneux

Millennium Challenge Corporation, Washington, DC, USA

The Millennium Challenge Corporation (MCC), a US foreign aid agency working with selected countries, emphasises rigorous evaluations to support poverty reduction through economic growth. MCC's experience with early impact evaluations and its growing evaluation portfolio has motivated actions to enhance the quality of evaluations. Specifically, MCC has introduced formal reviews that (i) better balance evaluation designs to ensure learning while respecting accountability, (ii) more selectively use impact evaluations and (iii) strengthen the programme logic and its documentation. MCC has also developed an explicit results dissemination strategy to ensure public access to evaluation results and most evaluation data, subject to ethical protections of respondents.

Keywords: learning; accountability; results measurement; development; impact evaluation; performance evaluation

1. A results-based agency

The Millennium Challenge Corporation (MCC) was established in 2004 as a US foreign aid agency that works in partnership with competitively selected countries that demonstrate a commitment to good governance, economic freedom and investment in their citizens. It was envisioned as an institution that would adhere to the key principles of the Paris Declaration on Aid Effectiveness. Once a country is selected as eligible for assistance, MCC works with the country to develop programmes focused on MCC's mission: reducing poverty through economic growth. MCC invests in two types of programmes: compact and threshold. Compacts are implemented by an accountable entity [typically referred to as the Millennium Challenge Account (MCA)] established by each partner government, while thresholds are implemented by MCC. As of June 2014, compact allocations comprised almost 95 per cent of total programme obligations (see [Table 1](#)).¹

In line with the fourth and fifth principles of the Paris Declaration, Results and Mutual Accountability, MCC's results framework provides a structure for implementing the agency's rigorous methods to project, track and evaluate the impacts of its programmes.²

This framework helps MCC answer basic aid effectiveness questions: Did the programme achieve its goals? Do the results justify the cost? While this paper focuses on MCC's independent evaluation activities, it is important to note the following regarding its overall results framework:

- As of June 2014, M&E allocations comprised 2 per cent of total programme obligations (see [Table 1](#)).³

*Corresponding author. Email: sturdyjr@mcc.gov

All authors are at the Millennium Challenge Corporation, www.mcc.gov.

Table 1. FY14Q3 MCC fund summary.

MCC fund sources	Sum of programme (excluding M&E) obligations	Sum of M&E obligations	Total programme obligations	Per cent of M&E to total programme obligations
Compact	\$8,874,560,926	\$140,542,898	\$9,015,103,824	2
Threshold	\$22,878,506	\$1,000,000	\$23,878,506	4
609(g)	\$161,370,548	\$4,776,007	\$166,146,555	3
Due diligence	\$285,219,978	\$64,778,742	\$349,998,720	19
Grand total	\$9,344,029,958	\$211,097,647	\$9,555,127,605	2

- Each programme has at least one MCC M&E staff member assigned to provide oversight of and support to M&E activities.
- Within each MCA, an M&E unit is established and responsible for the M&E Plan. The M&E Plan is a tool to manage the process of monitoring, evaluating and reporting progress towards results. It is used in conjunction with other tools such as work plans, procurement plans and financial plans.
- To monitor implementation and performance of activities during a compact's 5-year time frame, MCC uses a quarterly Indicator Tracking Table (ITT), generated by the MCA. These data are used for aggregate external reporting on common indicators across the MCC investment portfolio.
- For independent evaluation activities, MCC distinguishes two categories: impact and performance.⁴ As defined in MCC's M&E Policy, an impact evaluation is a study that measures the changes in income and/or other aspects of well-being that are attributable to a defined intervention. Impact evaluations require a credible and rigorously defined counterfactual, which estimates what would have happened to the beneficiaries absent the project. A performance evaluation is defined as a study where it is not warranted or possible to establish a credible counterfactual. Whether impact or performance, independent evaluations should include a process evaluation to document programme implementation in order to inform interpretation of results.
- Evaluations financed by MCC are managed by the Department of Policy and Evaluation. This department is independent from MCC's Department of Country Operations, which oversees the design and implementation of MCC's investments. Both departments report directly to the MCC's chief executive officer, and ultimately to the Executive Office of the President of the United States of America. This is unlike the independent evaluation offices of multilateral financial institutions which report to their respective governing boards.⁵
- MCC contracts external, independent organisations or individuals to design and implement the evaluations. While this approach is an essential element of MCC's commitment to unbiased, independent evaluations, it introduces particular challenges around how MCC enables the independence of the evaluations, as well as how the evaluators assert their independence. This issue is discussed in the final section of this paper.
- As part of MCC's commitment to transparency, all materials described above are made publicly available on its website, including M&E Plans, ITT data, evaluation materials and reports, and evaluation-related data, subject to practical and ethical protections of respondents.

Written 10 years into MCC's commitment to the Paris Declaration Principles, this paper reviews how MCC's implementation of this commitment has evolved. The following sections reflect on MCC's early evaluation practice, the lessons learned from the first five impact evaluations and how these, and other, lessons have informed MCC's thinking and practice around evaluation.

1.1. Early MCC evaluation practice (2004–2012)

Looking back at the brief history of evaluation at MCC, there are three characteristics worth highlighting: (i) an emphasis on impact evaluation, (ii) impact evaluations that prioritised accountability objectives to measure impacts on final outcomes and (iii) a degree of separation⁶ between MCC's M&E and programme operations that is characterised by independent lines of management and funding.

While there was no specific policy in place, there was an emphasis in MCC's early days to prioritise impact evaluations. This emphasis was a response to the lack of rigorous evidence available across development programmes and widespread calls for the need to contribute to this space (CGD 2006).

In addition to prioritising impact evaluations, MCC guided many of its early independent evaluators to prioritise accountability when conducting impact evaluations. That is, to answer the question: Did the programme improve key outcomes (particularly MCC's goal indicator – household income) as assumed in the original ex-ante economic rate of return? As a result, many evaluations designed during this time include a limited number of evaluation questions and outcomes focused on accountability measures, for example, income and its immediate determinants.

Last, there was a degree of separation between M&E and operations staff, influenced by the focus on accountability and a desire to enable independence of the evaluations. In practice, this degree of separation varied significantly between country, project and evaluation teams, and influenced how well evaluators understood the programmes they were evaluating and how involved sector teams were in defining evaluation questions, reviewing questionnaires and participating in other evaluation design-related discussions.

Given that the average MCC evaluation spans 5–7 years, from design through dissemination of results, it was not until 2012 with the release of the first five farmer training impact evaluations that MCC's approach to evaluation practice could be assessed across a number of evaluations in the same sector.

1.2. First five farmer training evaluations (2012)

The release of the first five farmer training impact evaluations offered internal and external stakeholders the opportunity to review several evaluations of one type of MCC investment to assess what was working and what wasn't. The results of these evaluations were shared publicly (<http://www.mcc.gov/pages/features/first-five-evaluations>) and are briefly summarised here.

These impact evaluations were designed to test whether results could be attributed to the MCC investments. While many targets were achieved according to monitoring data, the evaluations showed mixed results on income and are categorised as

- *Positive impacts on farm income.* In El Salvador, dairy farmers' annual farm incomes increased \$1849 over the control group (Blair et al. 2012).

- *Potentially positive, but further analysis required.* In Ghana, northern region farmers' annual crop income increased over the southern and central regions, though additional analysis would help to understand differences in impacts (ISSER 2012). In Nicaragua, an innovative evaluation technique suggests that annual farm household incomes increased \$2400, while standard methods suggest they did not. Additional analyses may explain the differences between the standard and innovative methods (Carter, Toledo, and Tjernström 2012).
- *No impacts on household income.* None of the evaluations listed above, or the Armenia evaluation (Fortson, Blair, and Gilbert 2012), with credible counterfactuals found convincing impacts on household income.⁷
- *Could not measure causal impacts.* In Honduras (NORC 2013) and the horticulture activity in El Salvador, the evaluations were not able to maintain appropriate control groups.

The core lessons that MCC operations and M&E took away from these results were the following:

Structure evaluations to facilitate learning. Like many evaluations designed during this time frame, the first five impact evaluations generally focused on measuring impacts on a few outcomes, particularly household income, and were not consistently structured to evaluate components of project design. For example, the farmer training programmes commonly included distribution of starter kits. However, kits were included as a package integrated with the farmer training. Consequently, the programmes and the evaluations were not designed to test whether the traditional assumption that starter kits increase adoption of new techniques was correct. In addition, several of these evaluations were not designed to assess impacts on intermediate outcomes *along* the causal chain to understand *why* the evaluators found particular results. The evaluations did not generally assess whether or not implementation went according to plan – How was farmer training defined? How many hours of training were required? Did complementary activities, in terms of credit access, value chain management, irrigation, occur? Understanding these factors could have contributed to understanding why an evaluator did or did not see expected impacts.

The Armenia evaluation was the exception that provided an example of the benefits of greater attention to exploring the causal chain. In this evaluation, the analyses of programme impacts on intermediate outcomes – specifically the effects on farming practices – provided important evidence that was used to challenge the validity of anomalous household income effects. The evaluation was also complemented by several qualitative studies designed to resolve questions about the effectiveness of the programme. This demonstrated how understanding *why* certain results were observed could provide more useful learning compared to analyses limited to a few accountability outcomes. In addition, this learning-oriented approach provides greater understanding of and confidence in the results that provide accountability.

Understand the programme design and implementation rules of the programme. Evaluation design *should* follow the programme design and implementation rules – more explicitly, the programme design and implementation rules need to be clear so

that M&E and evaluators know what they are evaluating. This came through in the following examples:

- In Honduras, the absence of a common understanding of how the implementer selected farmers for the training programme led to evaluation methodology assumptions that were never reconciled with the actual programme implementation. While the implementer initially confirmed farmers were selected on a set of quantified criteria,⁸ the implementer's contract was also performance-based, which created an incentive to find high-performing farmers who were able to meet the goals set in the contract. Meanwhile, M&E and the evaluator used the quantified criteria to construct an eligible pool of geographic areas to design a randomised control trial. Multiple attempts to construct a random sample of geographic areas that produced a large enough sample of farmers the implementer ultimately selected for the programme failed, in part due to the implementer's incentives to find high-performing farmers. This was not just a failure to have clear programme design and implementation rules, but also a failure to ensure that the design and implementation rules were well documented and clearly understood by all stakeholders and project participants.
- In Armenia, the primary agriculture investment renovated a failing irrigation system. The farmer training activity was added to train farmers on how to effectively use irrigation to produce higher-yield and higher-value field crops. However, the irrigation construction encountered major procurement-related delays, while the farmer training contracts were awarded close to the original timeline. Rather than pause the training programme, both the farmer training and its evaluation proceeded as planned. In retrospect, MCC's sector experts believe the necessary condition for the adoption of new techniques was the availability of a large-scale increase in irrigation water. Given the evaluation methodology was a randomised rollout, and the control groups received training before access to the improved irrigation, the randomised evaluation ultimately could not test the effect of the training after the irrigation became available. The lesson was for both implementers and evaluators to keep in mind the original programme logic to inform any necessary course corrections in both project and evaluation design and implementation.

The programme design and implementation rules that need to be considered for evaluation include (but are not limited to) (i) what is the unit of implementation – at what level will the programme be implemented (village, district)? Understanding the unit of implementation and how the programme will be managed from that unit may inform the unit of assignment in a randomised control trial. (ii) What is the selection criteria for the programme area? For programme participants? This information is not only useful to help determine how a credible counterfactual may be constructed, but also to understand the potential external validity of results. (iii) Is the programme part of a package of complementary activities? Do results depend on coordination and implementation of all these activities? The feasibility and credibility of the evaluation methodology is related to how these factors are considered throughout design and implementation.

Build in mechanisms to assess evaluation risks, feasibility, costs and benefits. It is often the case that programme design and implementation rules evolve and change shape over time. Even if an impact evaluation design is reflective of programme design and

implementation rules early in the programme, this may not be the case as the programme evolves. For example, the use of randomised roll-out approaches in four of the five evaluations seemed feasible at baseline, when the programme implementation schedule allowed for the necessary gap between the start of the treatment and the start of the control trainings to guarantee a sufficient exposure period for impacts on outcomes. However, in several cases and for a variety of reasons, that exposure period was reduced.⁹ Given that course corrections and programme re-scoping commonly occur in programmes, evaluation methodologies, whether using pure control groups, comparing multiple interventions or using a standard pre-post methodology, should be regularly assessed to determine whether they remain feasible and informative. In order for this to occur, both processes and incentives must be in place to get the right people around the table at the right time to make these assessments.

Ensure timing of evaluations follow the programme logic and assumptions. There is often pressure to demonstrate results quickly, perhaps to inform decisions for scale-up or to support similar investments. In the case of impact evaluations, there is also pressure to incorporate or compensate control groups. However, stakeholders should recognise that sufficient exposure to treatment, based on a well-conceived programme logic, is required for the evaluation to measure changes in outcomes. For example, the Nicaragua results suggest that the major benefits of that programme did not start accruing until the second and third years of the programme. However, in cases like Armenia and Ghana, it was noted prior to and during implementation that 12-month exposure would be sufficient. Subsequently, peer reviewers of the final evaluation results speculated these evaluation periods were likely inadequate time for farmers to learn and confidently apply new methods and see increases in income. In order to mitigate this sort of challenge to evaluation results in the future, exposure periods should be based on the programme logic and specified and assessed from the beginning. This should then inform data collection timelines, as well as expectations around how long control groups need to be maintained and when results will be available. In addition, a rigorous monitoring system would support timely reporting on output- and outcome-related data that could inform whether or not the exposure period hypothesised in the programme logic was sufficient and the data collection plan remains valid.

Evaluators and implementers must work in lock-step. In order to turn the first four lessons learned into practice, evaluators and implementers must work closely together. Collaboration between the teams throughout design and implementation is critical to ensure selection of beneficiaries, sequencing and exposure periods are agreed upon and carried out with a shared commitment to successful and effective execution of both the project and evaluation. In addition, close integration between evaluators and implementers can help identify potential users of evaluation results. By working closely to identify and answer policy-relevant questions posed by implementers and other key stakeholders, the evaluations should produce learning directly applicable to future programme selection, design and implementation. The line between *independent* evaluator and implementer should always exist, as a critical factor to produce unbiased evaluation results; however, this should never obstruct a clear and shared understanding of both the programme and its evaluation.

1.3. Adapting MCC thinking about evaluation (2012–2014)

According to MCC's current¹⁰ evaluation portfolio, there are almost 150 completed, planned or ongoing independent evaluations, 40 per cent of which are classified as impact and 60 per cent of which are performance. Based on the lessons described above, and another 21 completed MCC evaluations publicly available, MCC is adapting its thinking around this evaluation portfolio in order to (i) *be selective in how and when to use impact evaluation*, and (ii) *find a better balance between accountability and learning*.

Be selective in how and when to use impact evaluation. Well-designed and executed impact evaluations are powerful tools for generating evidence regarding development interventions. While the former training impact evaluations summarised above highlight challenges with both project and impact evaluation design and implementation, other impact evaluations in MCC's portfolio provide evidence for MCC and partner governments on what works and what doesn't. Findings in Burkina Faso demonstrated that new primary school infrastructure led to increased enrolment rates and improvements on math and French test scores (Levy et al. 2009); in Niger, investments in new primary school infrastructure led to increases in school enrolment, but no impact on attendance or on math and French test scores, although impacts were generally larger for girls than for boys (Dumitrescu et al. 2011); in Mozambique, installation of rural water pumps increased household access to improved water and reduced time spent fetching water, yet did not have impacts on overall quality of water stored at the household or health outcomes (Hall et al. 2014). The viability of these evaluations lay in three factors: (i) they measured outcomes along the causal chain, (ii) outcomes were measured in a time frame that was reasonable given the exposure period and (iii) the evaluation designs provided credible counterfactuals.

While MCC recognises the potential usefulness of well-designed and implemented impact evaluations, the learning from the first five impact evaluations demonstrated just how hard it is to design and execute programmes with integrated, rigorous impact evaluations. With this in mind, MCC is adapting its approach from one that prioritises impact evaluation to one that is selective in how and when to use impact evaluation, and when a performance evaluation is more appropriate. In being selective, the first condition for doing an impact evaluation is to assess the feasibility of establishing a credible counterfactual group, either through random assignment or other appropriate techniques, which is consistent with programme design and implementation requirements. This assessment requires understanding the rules of project design and implementation that can enable, or prevent, estimating a counterfactual.

Once it is understood that a counterfactual is feasible, there are additional factors that may weigh in favour of conducting an impact evaluation, including (i) there is demand by stakeholders, particularly local government for rigorous evaluation results, and therefore commitment to maintain control groups; (ii) there is learning potential, including potential for the project to be expanded in the country or other similar contexts; and (iii) there are significant gaps in evidence about what works and what doesn't. Factors that weigh against conducting an impact evaluation include (i) stakeholder resistance to the evaluation methodology and (ii) where time and financial costs outweigh the expected benefits of the evaluation. For programmes where factors ultimately weigh against doing an impact evaluation, MCC is committed to strengthening how it uses performance evaluations and other results measurement activities to meet accountability and learning objectives.¹¹

Balance accountability and learning. Another balancing act is between the accountability and learning¹² objectives of the M&E activities, especially impact evaluations. The accountability objective can be viewed from at least two levels. One level relates to outputs: What did we intend to buy and did we get what we paid for? There are a variety of programme and M&E activities that can help inform this objective, from implementer field reports, to monitoring data, to an independent evaluator's process evaluation. While all of these sources are possible, what became clear as a result of the first five impact evaluations was that understanding the programme – what was intended and what was realised in terms of outputs – is a fundamental component to understanding and interpreting results on outcomes.¹³

A second level to the accountability objective relates to the ultimate outcomes – Did the programme demonstrate adequate levels of the intended results? At MCC, this has centred on – Did household income increase sufficiently to warrant the programme investment? The first five farmer training evaluations did not detect an impact on household incomes, despite meeting or exceeding monitoring targets in all five and detecting increases in farm incomes in three, which raises two interesting points. First, we may not know enough to predict the time pattern of total household income impacts of the programme, particularly if income has multiple sources, and beneficiaries reallocate labour across sources. Second, increases in outcomes like farm income, and subsequently total household income, are prone to gestational lags in learning, adopting and benefitting economically from new methods. With this in mind, the evaluations demonstrated just how difficult it is to measure income, how critical assumptions about timing are and how those assumptions affect what outcomes can and should be measured in evaluations.

MCC makes investment decisions based on ex-ante economic rates of return that typically have 20-year time horizons. Investments are implemented during a 5-year period, with implementation sometimes completing by the end or even after the 5-year compact. Evaluation timelines need a certain amount of exposure to treatment once implementation has completed in order to demonstrate results. The evaluation then requires data collection, analysis and write-up of results. As discussed above in lessons learned, this timeline often receives pressure to respond to requests to demonstrate results quickly, as well as incorporate or compensate control groups.

These pressures can reduce the ability to rigorously measure the long-term results from investments. If stakeholders call for measurement of immediate impacts or if control populations need to be incorporated into the treatment group, the evaluation time frame may not allow for an adequate exposure period. For this reason, it may be unrealistic to assume household income could increase in the evaluation time frame. MCC's early evaluation experience therefore suggests the need to increase the emphasis on what is measured and assessed along the causal chain to understand whether or not the investment led to improvements in immediate and medium-term outcomes.¹⁴

In addition, while recognising the challenges in measurement, MCC still aims for its evaluations to measure longer-term impacts, particularly household income as the best measure of poverty reduction. For programmes that have a reasonable expectation of increasing household income in the evaluation time frame, household income remains the primary outcome of interest. For programmes where the evaluation time frame may not cover longer-term outcomes, including household income, in addition to measuring outcomes along the causal chain, MCC looks for opportunities to enable follow-up studies that can explore long-run impacts of programmes (see Bagby et al. 2013; Sloan and Levy 2012). In addition, by providing public access to its evaluations' methodologies and data, MCC enables not only its evaluators but also outside researchers to explore possibilities of

revisiting evaluation samples years later to measure longer-term impacts, particularly when control groups remain viable.

2. Challenges

There are common challenges facing agencies like MCC as they work towards maximising the value of investments in results measurement, particularly in impact evaluation.

2.1. *Early integration of evaluators with implementers vs. project readiness*

MCC often faces tensions over bringing evaluators into discussions on programme design early enough to initiate discussions on how to monitor and evaluate, vs. bringing them in too early – before a programme is sufficiently designed so it is clear what will be monitored and evaluated. This can lead to potentially costly revisions in evaluation design as it adapts to an evolving programme logic. It also risks turning the evaluator into a full co-author of the evolving programme design. While this might strengthen programme design in some cases, it also jeopardises the credibility of the evaluator's independent assessment.

2.2. *Timely assessments of design and implementation risks*

The quality of any evaluation portfolio is affected by a range of factors: quality of project design and readiness, feasibility of evaluation methodology, commitment of stakeholders to the evaluation methodology, appropriate definition of outcomes and corresponding questionnaires for measurement, adherence to agreed roll-out plan for treatment and control groups, external validity of results, internal validity of evaluation sample, sufficient sample sizes and so on. At any time, threats to any of these factors can derail an evaluation. Getting the right people around the table at the right times to assess risk and make decisions on how to course correct when necessary is usually difficult given competing priorities, workloads and varying technical and managerial skills and responsibilities. Tackling this challenge requires better alignment of incentives for stakeholders – M&E, management, sector leads, country leads and evaluators – to engage in timely and quality assessment of evaluation designs and products.

2.3. *Appropriate documentation of decisions*

As with many other organisations, MCC deals with turnover in M&E, operations, government counterparts, implementer staff and evaluator staff. Establishing a mechanism for central documentation of key decisions made on both project and evaluation design and implementation issues is critical to addressing staff turnover and enabling consistent management of evaluations over the many years from development to completion.

2.4. *Timely dissemination and use of results*

Establishing a strong feedback loop – getting the results to the right people at the right time to make informed decisions – is a challenge. Because MCC compacts are limited to 5 years, many evaluation results come out after the MCC's programmatic relationship with a country ends and the original country teams have dissolved. This problem is not unique to MCC; aligning the need for evidence with the availability of evidence is always a challenge.

3. Turning lessons into action

Considering the lessons learned and challenges described above, the following summarises the key actions implemented by the MCC M&E division.

3.1. Use evaluability assessment to inform evaluation decisions

Evaluability is defined as the ability of an intervention to demonstrate in measurable terms the results it intends to deliver (IDB 2010). An evaluable intervention uses data to identify and verify the problem(s) it intends to address, has an evidence-based design to address the problem(s) identified, identifies assumptions and risks associated with the intervention to address the problem(s), including means for verifying and mitigating risks, and has clear and time-bound metrics for output and outcome results.

MCC M&E has developed and is applying an evaluability assessment tool to use specific, transparent standards and best practices for assessing the five dimensions of a project described above based on project documentation. Working towards designing and implementing evaluable interventions is intended to strengthen project design and implementation rules that affect evaluation. The assessment is intended to determine whether or not projects are evaluable; a necessary condition to warrant an evaluation.

3.2. Establish a formal review process for technical and financial decisions

MCC established an Evaluation Review and Management Process in 2013 with the following objectives:

- *Enable independence of evaluations and allow evaluators to assert independence.* The review process introduces structure for how MCC manages independent evaluations and provides a framework that allows evaluators to assert their independence. MCC and MCAs play a critical role in defining evaluation questions and assessing feasibility of proposed evaluation methodology. However, once the evaluation design is approved, MCC's role in the evaluation implementation process shifts. Evaluators are completely responsible for findings and reports. During the review process for final reports, MCC and MCA's role is limited to providing comments on the technical and factual accuracy of the reports. All comments are documented and made public as an annex to the evaluator's report for full transparency of the review process.
- *Identify stakeholders and establish feedback loops.* An internal Evaluation Management Committee (EMC) is now established for each compact, chaired by the MCC M&E managing director, or a designated representative, and consisting of the appropriate representation from M&E and operations, including senior management when appropriate. Country stakeholders, such as the MCA, government, implementing entities and MCA contractors, are identified and included as necessary in each stage of the evaluation process. In addition, MCAs are encouraged to establish their own, parallel EMC to facilitate appropriate local stakeholder review of the materials.
- *Identify milestones.* Milestones have been established to review evaluation tasks: (i) the evaluation plan and the evaluator scope of work, (ii) the evaluation design and subsequent evaluation deliverables (survey instruments, baseline report) and (iii) the interim/final evaluation report(s). In addition to these milestones, the EMC meets

throughout the life of the evaluation to assess course corrections, particularly those that require a change in time and financial resources dedicated to the evaluation. To assist with assessment of evaluation activities and products, MCC has created a set of templates, including an evaluator scope of work and an evaluation risk assessment checklist.

- *Find the right balance for accountability and learning.* By identifying the right stakeholders, both internal and external, the review process is the mechanism for balancing the trade-offs between accountability and learning. There is no one-size-fits-all approach to evaluation, it depends on what is being evaluated whether the evaluation will lean more towards accountability – did it do what it said it would – or learning. The important action is to create a mechanism for this decision to be made with all appropriate stakeholders and documented. The EMC is responsible for assessing what result measurement tools, including impact, performance evaluation and other monitoring activities, are available in order to meet accountability and learning objectives.
- *Document decisions.* Given the considerable time and financial resources committed, it is critical for management and staff to document the key evaluation design and implementation decisions. The Evaluation Review Process creates a forum for who makes what decisions when and enables formal documentation of that process.

As MCC learned during the release of the first five evaluations, its evaluations need to be able to stand up to internal and external assessment, including questions around whether sample sizes were sufficient to believe results; whether the timing of surveys was appropriate; whether the evaluators were measuring the right outcomes in the right way; and whether evaluators, as hired consultants, remain unbiased in their assessments as opposed to catering to their client. The formal review process, assuming appropriate participation by key stakeholders, is viewed as one of MCC's key mechanisms for mitigating these risks and ensuring investment in high-quality evaluations.

3.3. Define the results dissemination strategy

Ultimately, investments in the evaluation portfolio are only as good as the extent to which results are used for current and future decision making. As a result, MCC has defined a results dissemination strategy, which also establishes a strong internal and external feedback loop, including the following:

- *MCC Evaluation Catalog* (<http://data.mcc.gov/evaluations/index.php/catalog>). In collaboration with the World Bank's Development Data Group, the MCC Evaluation Catalog was launched in 2013 as the primary mechanism for disseminating the primary evaluation materials associated with its evaluations (evaluation design reports, questionnaires, baseline reports, analysis reports and data).
- *MCC Summary of Findings.* The Summary of Findings is an MCC-authored document 5–6 pages in length that summarises the evaluated programme in context of the overall compact, the programme logic, assumptions, monitoring indicators and results, as well as evaluation questions and key results. The Summary of Findings also summarises key lessons learned by MCC that directly result from the evaluation findings.
- *MCC Management and Country Statements.* The MCC Management Statement, as well as the Country Statement, confirms acceptance of the final report and

documents any outstanding differences between MCC, or the partner country, and the Evaluation Firm relating to (i) factual and/or (ii) technical issues.

3.4. Establish evaluation data protection practices

MCC is committed to publicly sharing data generated in the design, implementation, monitoring and evaluation of its programmes to achieve two important objectives:

- *Transparency.* To allow any stakeholder access to the data and analysis behind MCC evaluations to enable validation of, or challenge to, results.
- *Policy research for the public good.* Public access to MCC-financed data can stimulate a wide range of policy-relevant research, maximising the benefits of MCC's investments in large-scale data collection efforts in developing countries.

While all monitoring data are available at an aggregate level that poses low risk of harm to any individual, the data collected at the community, household and individual levels for independent evaluations must be collected, stored and disseminated in a way that not only meets the above objectives but meets them in a way that minimises risk of harm to the respondents. With this in mind, MCC has put in place several data protection practices:

- *Institutional Review Board (IRB) requirement.* MCC M&E introduced a contractual requirement for independent evaluators to submit research protocols detailing the evaluation design and methodology, as well as data collection, storage and dissemination proposal, to an accredited IRB. The objective of this requirement is to ensure that the independent evaluator, as well as any data collection firm under their direct or indirect management, establishes appropriate procedures for protecting survey participants' confidentiality and minimising risk of harm to the participants as a result of their participation in the evaluation of data collection.
- *Data anonymisation and review process.* In order to publicly release data, evaluators must conduct data anonymisation to prevent re-identification of survey participants. Once completed, the anonymised data are submitted to the MCC Disclosure Review Board (DRB). The DRB was established in June 2013 to review public use data files with the objective of balancing between maximising use of the data and minimising risk of harm to survey participants. The DRB documents decisions on releasing, or not releasing data, based on the review of anonymisation procedures. Once a public use data file is cleared for dissemination by the DRB, it is posted on the MCC Evaluation Catalog.

Just as MCC has adapted its evaluation practices as a result of the challenges and lessons learned in the first five farmer training impact evaluations, other well-designed impact evaluations are also providing evidence that influences MCC's new programmes. By documenting the limited impacts that school infrastructure investments had on learning outcomes in Niger, the evaluation findings raised a question that influenced MCC's decision to invest in soft interventions: Could these soft interventions improve learning outcomes more than school construction alone?¹⁵ In a similar vein, as noted earlier, impacts of improved infrastructure alone on water-related health outcomes were limited in Mozambique, so MCC is working to integrate its large water infrastructure investments

with hygiene and sanitation behaviour change interventions in new projects like those in the Zambia Compact¹⁶ to see whether we can increase their net benefits.

The efforts described above were taken to strengthen our evaluation practices in order to improve and increase evidence informing future investments. These efforts reflect our commitment to the principle that the benefits of strategic investment in well-designed and implemented impact evaluations can and must outweigh their costs.

Acknowledgements

The authors wish to acknowledge this paper would not have been possible without the collaboration with Kathy Farley, Kristin Penn and Sarah Lucas on the lessons learned from the first five farmer training evaluations; the contributions of Olivier Dupriez, Mehmood Asghar, Heather Hanson, Nicole Styler and Paul Burgess to support development of MCC's Evaluation Catalog for data dissemination practices; as well as feedback from the participants in the June 2014 Norad-3ie writing seminar. The authors would also like to thank Kartik Akileswaran, Timothy Breitbarth, Scott Fontaine, Jed Friedman, Rebecca Goldsmith, Sheila Herrling, Sebastian Martinez, Ryan Moore and Sarah Jane Staats for comments on earlier drafts.

Funding

All evaluations reviewed in this paper were funded by MCC.

Disclosure statement

The authors are employees of the Millennium Challenge Corporation (MCC). The evaluations are publicly available on MCC's website, www.mcc.gov. The opinions expressed in this paper are those of the authors and do not necessarily reflect the view of MCC or the US government. The authors declare that other than the compensation they receive as MCC employees, they have no financial interest in the research described in this paper.

Notes

1. As indicated in Table 1, there are four main fund sources: compact, threshold, 609(g) and due diligence.
2. For more detail, please refer to the 2012 MCC Policy for Monitoring and Evaluation of Compacts and Threshold Programs – <http://www.mcc.gov/documents/guidance/policy-050112-monitoring-and-evaluation.pdf>.
3. These are projected costs for any ongoing programme.
4. For more detail, see the MCC Independent Evaluations Issue Brief – <http://www.mcc.gov/pages/docs/doc/issue-brief-mcc-independent-evaluations>.
5. For example, the World Bank, the International Monetary Fund, the Inter-American Development Bank, African Development Bank and the Asian Development Bank all have independent evaluation offices that were authorised by and report to their respective governing boards.
6. The organisational structure of MCC shifted several times over these 8 years, with M&E sometimes separated from and sometimes embedded in the operations division.
7. The 'convincing' qualifier is added here because one regional subset of results in the Armenia evaluation indicated a marginally significant programme impact on household incomes. But the evaluator assessed this result as unlikely to reflect a true programme effect because the regional results did not indicate any significantly greater programme impact on farming practices – which was the essential intermediate outcome through which the programme was intended to affect incomes.
8. The criteria included access to water, less than 50 hectares of land, interest in adopting programme and not receiving other technical assistance.

9. For Ghana, the original exposure period was 24 months. However, implementation took longer than expected in the first phase, and sponsors of the training and evaluation agreed to shorten the lag time between groups to 1 year due to pressure to scale up in controls. MiDA and agriculture specialists thought farmers would be able to show changed behaviour and gains in yields after 1 year, although the changes were expected to be smaller than originally anticipated. M&E and the evaluators thought those smaller changes could be measured by doubling the sample. In El Salvador horticulture, the project underwent a re-design in the middle of the evaluation period, resulting in the control group being incorporated into the training after only a very short exposure period between treatment and control. The randomised roll-out approach was successfully implemented in Armenia, El Salvador dairy and El Salvador handicrafts.
10. As of the FY14Q3 Evaluation Pipeline Summary.
11. For an example of a recently designed and completed MCC-financed performance evaluation, see Mongolia Health Project (de Graaf 2014).
12. MCC refers to ‘accountability’ and ‘learning’ as two different objectives of evaluations, This difference, while useful, might suggest that they are distinct. But as Markus Goldstein, in his insightful blog, ‘The Tao of Impact Evaluation’, clarifies, the two objectives are not orthogonal; learning itself can contribute to accountability. (See <http://blogs.worldbank.org/impactvaluations/the-tao-of-impact-evaluation>.)
13. For an example of an ongoing MCC-financed evaluation designed to measure project process and implementation to explain impacts, see Philippines KALAH-CIDDS (IPA 2011).
14. For an example of an ongoing MCC-financed evaluation designed to measure primary outcomes along the causal chain, recognising the limitations of what can be measured during the evaluation period, see Tanzania Water Project (Duthie, Alwang, and Pendley 2012).
15. For more information, see Plan 2012 (<http://www.planusa.org/content3120321>).
16. For more information, see MCC 2012 (<http://www.mcc.gov/documents/agreements/compact-zambia.pdf>).

References

- Bagby, E., E. Borkum, A. Dumitrescu, and M. Sloan. 2013. *Evaluation of the Niger Education and Community Strengthening Program Design Report*. Mathematica Policy Research.
- Blair, R., L. Campuzano, L. Moreno, and S. Morgan. 2012. *Impact Evaluation Findings after One Year of the Productive and Business Services Activity of the Productive Development Project, El Salvador*. Mathematica Policy Research. http://data.mcc.gov/evaluations/index.php/catalog/63/related_materials
- Carter, M., P. Toledo, and E. Tjernström. 2012. *The Impact of Rural Business Services on the Economic Well-being of Small Farmers in Nicaragua*. http://data.mcc.gov/evaluations/index.php/catalog/74/related_materials
- CGD (Center for Global Development). 2006. *When Will We Ever Learn? Improving Lives through Impact Evaluation*. <http://www.cgdev.org/publication/when-will-we-ever-learn-improving-lives-through-impact-evaluation>
- de Graaf, P. 2014. *Health Project MCA-Mongolia Non-Communicable Diseases and Injuries – Evaluation Final Report*. http://data.mcc.gov/evaluations/index.php/catalog/97/related_materials
- Dumitrescu, A., D. Levy, C. Orfield, and M. Sloan. 2011. *Impact Evaluation of Niger’s IMAGINE Program Final Report*. Mathematica Policy Research. http://www.mathematica-mpr.com/~media/publications/pdfs/international/niger_imagine.pdf
- Duthie, M., J. Alwang, and C. Pendley. 2012. *Impact Evaluation Design Report MCA Tanzania Water Sector Project*. Social Impact. http://data.mcc.gov/evaluations/index.php/catalog/86/related_materials
- Fortson, R., L. Blair, and Gilbert. 2012. *Evaluation of Water-to-Market Training in Armenia*. Mathematica Policy Research. http://data.mcc.gov/evaluations/index.php/catalog/56/related_materials
- Hall, R., J. Davis, E. Van Houweling, E. Vance, M. Carzolio, M. Seiss, and K. Russel. 2014. *Impact Evaluation of the Mozambique Rural Water Supply Activity Under a Cooperative Agreement between MCC and Stanford University*. http://data.mcc.gov/evaluations/index.php/catalog/76/related_materials
- Innovations for Poverty Action. 2011. *MCC KALAH-CIDDS Evaluation: Design Report*. http://data.mcc.gov/evaluations/index.php/catalog/59/related_materials

- Institute of Statistical, Social and Economic Research, University of Ghana. 2012. *An Impact Evaluation of the MIDA FBO Training*. http://data.mcc.gov/evaluations/index.php/catalog/77/related_materials
- Interamerican Development Bank. 2010. *Evaluability Review of Bank Projects 2009*. Office of Evaluation and Oversight, OVE. <http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=35594535>
- Levy, D., M. Sloan, L. Linden, and H. Kazianga. 2009. *Impact Evaluation of Burkina Faso's BRIGHT Program Final Report*. Mathematica Policy Research. http://www.mathematica-mpr.com/~media/publications/pdfs/international/burkina_bright.pdf
- NORC at the University of Chicago. 2013. *Impact Evaluation of the Farmer Training and Development Activity in Honduras*. http://data.mcc.gov/evaluations/index.php/catalog/52/related_materials
- Sloan, M., and D. Levy. 2012. *BRIGHT II Evaluation Design (Revised)*. Mathematica Policy Research. http://data.mcc.gov/evaluations/index.php/catalog/92/related_materials